

Data Archive

The SURF Data Archive allows users to safely archive up to petabytes of valuable research data to ensure the long term accessibility and reproducibility of their work. The Data Archive uses tape technology to provide affordable, safe, and secure data storage. The Data Archive is also connected to SURF's compute infrastructure, via a fast network connection, allowing for the seamless depositing and retrieval of data.

Who is it for?

The Data Archive is available to all SURF members. Access can be granted through several streams depending on the user's affiliation and institutional contracts.

Affiliate of SURF Member Institution – Using SURF Computing Services

If you are a researcher at a [SURF member institution](#), you can submit an individual request to SURF for Data Archive access as part of the application process for the computing services (ex. Snellius and Research Cloud). For more details see the [access to the computing services page](#).

If you are a researcher at a SURF member institution with a dedicated contract for Data Archive use across the whole institution, an individual request does not need to be submitted and capacity will be available on demand. This type of contract is negotiated between SURF and the institution's IT service. To find out if your institution has this type of contract and the stipulations of use, please contact your local IT service department.

Affiliate of SURF Member Institution – Require a Dedicated Contract for Research Group/Project

If you are a researcher at a SURF member institution who requires dedicated space on the Data Archive for a project or research group and do not require any other SURF computing services, you can request a quote based on the amount of space and duration length required via servicedesk@surf.nl or our service portal.

Affiliate of Educational/Research Facility or Private Business Enterprise Not Associated with SURF

If you are a member of an organisation that is not a SURF member but would like to take advantage of our Data Archive services, please email us at servicedesk@surf.nl. Our advisors will determine if you are eligible to use the Data Archive and can provide additional information about the applicable rates.

Why should I archive my data?

There are 2 main reasons to archive data: scientific integrity and data reuse.

Scientific Integrity

The [Netherlands Code of Conduct for Research Integrity](#) has been adopted by the NWO, KNAW, NFWO, VSNU and TO2 Federation. It is based on a set of principles: honesty, scrupulousness, transparency, independence, and responsibility. By retaining all the data required to recreate the research done, researchers ensure that their work is transparent and honest.

Data Reuse

Equally important to transparency and honesty in scientific research is maximizing the value of research. Researchers can store datasets that may prove useful in future research endeavours (either through novel analyses or aggregation with other datasets). This helps researchers to [FAIRify](#) their data and reduces the investment cost of future research (as long-term cold storage is usually cheaper than repeating the experiment).

What data can I store?

The Data Archive is aimed specifically at the secure storage of large volumes of data that are not actively in use. For instance, researchers can opt to 'freeze' data from an article or store raw data which may be reused in some yet unknown future research. The Data Archive is not meant for keeping an incremental copy of data/directories in case the primary copy is lost (aka backups) or as a cheaper option for hosting large datasets that are undergoing regular access or processing on one of our computing services. For more details about effective data storage and the effects on the system see [Effective Archive File Management](#).

Table of contents

- Who is it for?
 - Affiliate of SURF Member Institution – Using SURF Computing Services
 - Affiliate of SURF Member Institution – Require a Dedicated Contract for Research Group /Project
 - Affiliate of Educational /Research Facility or Private Business Enterprise Not Associated with SURF
- Why should I archive my data?
 - Scientific Integrity
 - Data Reuse
- What data can I store?
 - Guidelines
 - Do Store
 - Do Not Store
- Where is my data stored?
- How can I start using it?

Guidelines

- Files should be between 1GB and 200GB in size and average file size should not fall below 1GB.
 - Details on how to pack directories of small files into larger files or divide extra large files into appropriately sized chunks can be found [here](#).
- Files should be packed before uploading them to the archive.
 - We recommend using `dmftar` as detailed [here](#).
- Plan out how data is spread across the file and folder structure such that specific data will be easy to find and restore without recalling unneeded data.
 - Documenting the contents of files and folders with a text file is also strongly recommended – even a couple words per file can save a lot of frustration down the road.

Do Store

- Data from completed research projects
 - Raw data like unprocessed audio files, field notes or readings straight from machine equipment
 - Processed data like digitized drawings, transcribed interviews, validated datasets and anonymized survey results
 - Analyzed data like models, graphs, tables and texts that convey useful information, decisions or conclusions
- Key temporal snapshots of datasets processed on one of the compute services
- Identity files for authentication and authorization
- Configuration files for applications and tools used on the archive

Do Not Store

- Unpacked software packages
 - Packed software packages are acceptable
- Backups of working directories (e.g. Snellius project space)
- Computer/phone/peripherals backups
- Personal data not related to your research output (e.g. your music library, photo collections, tax returns, etc)
- Unpacked research datasets

If you have any questions about whether your use case is appropriate for the Data Archive or need advice on how to format it for optimal storage, please contact our advisors at servicedesk@surf.nl or [our service desk portal](#). An [Acceptable Use Policy](#) is also available.

Where is my data stored?

The Data Archive maintains two tape libraries for security and redundancy in two physically separate locations in the Amsterdam and Haarlemmermeer municipalities. When data is uploaded to the Data Archive using SSH, (HPN)SCP, SFTP, rsync, GridFTP, iRODS, etc. it ends up on an online disk space managed by the Data Migration Facility (DMF). The DMF will then manage the careful migration of files from the disk space to two tape libraries until your data is available on both tape libraries. Once your data is safely stored in the two tape libraries it may be removed from the disk space (aka offline). Offline data can be interacted with in the same manner as online data though users may notice a delay in access time.

How can I start using it?

To set up an account you can contact us via servicedesk@surf.nl or [the service portal](#). Any questions or special requests can also be directed there.